Formulas

This is a complete collection of all formulas used in the Computing Infrastructures exercises.

Dependability

Reliability

The probability that a system will operate correctly until time t:

$$R(t) = P(T \geq t) = e^{-\lambda t},\ \lambda = \frac{1}{\text{MTTF}} \text{ failure rate}$$

Availability

The probability that a system will be operational at time t:

$$A(t) = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}}$$

Mean Time Between Failures

The average time between to distinct failures of a system:

$$\text{MTBF} = \frac{\text{total operating time}}{\text{number of failures}}$$

Failures in Time

The number of expected failures per one billion hours of operation:

$$\text{FIT} = \frac{10^9}{\text{MTBF}}$$

Reliability block diagrams

Blocks in series For blocks in series, the total reliability is:

$$R(t) = \prod_{i=1}^{n} R_i(t)$$

The total availability, instead, is:

$$A(t) = \prod_{i=1}^{n} A_i(t)$$

Blocks in parallel For blocks in parallel, the total reliability is:

$$R(t) = 1 - \prod_{i=1}^{n} \left[1 - R_i(t)\right]$$

The total availability, instead, is:

$$A(t) = 1 - \prod_{i=1}^{n} \left[1 - A_i(t)\right]$$

Standby redundancy

We define a system to be “standby redundant” if it is made up of two parallel replicas where a redundant replica is activated via a switch if the primary replica fails.

Perfect switching If the switch in the system never fails—that is, it has a reliability of 1—if the two blocks in the system have the same failure rate λ, then:

$$R(t) = e^{-\lambda t}(1 + \lambda t)$$

While if the two blocks have failure rates $\lambda_1 \neq \lambda_2$:

$$R(t) = e^{-\lambda_1 t} + \frac{\lambda_1}{\lambda_2 - \lambda_1}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right)$$

Imperfect switching If the switch can fail—so its reliability is $R_{\text{switch}} < 1$—we must take it into account as well.

If the two blocks have identical failure rates:

$$R(t) = e^{-\lambda t}(1 + R_{\text{switch}}\lambda t)$$

While if the two blocks have failure rates $\lambda_1 \neq \lambda_2$:

$$R(t) = e^{-\lambda_1 t} + R_{\text{switch}}\frac{\lambda_1}{\lambda_2 - \lambda_1}\left(e^{-\lambda_1 t} - e^{-\lambda_2 t}\right)$$

n **redundant replicas** In general, for one primary replica with n redundant replicas in parallel, if the switch never fails, the total system reliability is:

$$R(t) = e^{-\lambda t}\sum_{i=0}^{n-1}\frac{(\lambda t)^i}{i!}$$

R-out-of-N redundancy

We define a system as "R-out-of-N" redundant if it is made up of N identical replicas where at least R replicas must be functioning in order for the system to be online. The functioning replicas are identified via a voter component, which has its own reliability R_{voter}.

The total reliability of a system like this is:

$$R(t) = R_{\text{voter}}\sum_{i=R}^{N} R_{\text{component}}^{i}(1 - R_{\text{component}})^{N-i}\binom{N}{i}$$

HDDs

In order to compute the complete transfer time of a Hard Disk Drive, we can calculate:

$$T_{\text{I/O}} = (1 - \text{DL})(T_{\text{seek}} + T_{\text{rotation}}) + T_{\text{transfer}} + T_{\text{controller}}$$

where:

$$T_{\text{rotation}} = 60\frac{s}{\text{min}} \cdot \frac{1}{2 \cdot \text{Rotation speed in RPM}}$$

and:

$$T_{\text{transfer}} = \frac{\text{Block size}}{\text{Transfer speed}}$$

RAID

Below, a table defining all capacities and MTTFs of RAID configurations. N is the number of disks in the RAID system, while C is the capacity of a single disk.

RAID type	Capacity	MTTF
0	NC	$\frac{\text{MTTF}_{\text{Disk}}}{N}$
1	C	$\frac{\text{MTTF}^2_{\text{Disk}}}{N \cdot \text{MTTR}}$
0+1	$\frac{N}{2} \cdot C$	$\frac{\text{MTTF}^2_{\text{Disk}}}{N \cdot G \cdot \text{MTTR}}$
1+0	$\frac{N}{2} \cdot C$	$\frac{\text{MTTF}^2_{\text{Disk}}}{N \cdot \text{MTTR}}$
4	$(N-1)C$	$\frac{\text{MTTF}^2_{\text{Disk}}}{N(N-1)\text{MTTR}}$
5	$(N-1)C$	$\frac{\text{MTTF}^2_{\text{Disk}}}{N(N-1)\text{MTTR}}$
6	$(N-2)C$	$\frac{\text{MTTF}^2_{\text{Disk}}}{N(N-1)(N-2)\text{MTTR}}$

Performance evaluation

Parameters

Below is a list of parameters used in performance evaluation:

Parameter	Description
T	Observation time
N	Population of a system
C	Number of completed jobs in a system
Z	User think time (only for terminal systems)
A_k	Arrivals at resource k
C_k	Completions at resource k
B_k	Busy time of resource k
N_k	Population at resource k
D_k	Demand of resource k
V_k	Visits at resource k

Operational laws

Arrival rate

$$\lambda_k = \frac{A_k}{T}$$

Throughput

$$X = \frac{C}{T}$$

Utilisation

$$U_k = \frac{B_k}{T}$$

Service time

$$S_k = \frac{B_k}{C_k}$$

Visits

$$V_k = \frac{C_k}{C}$$

Demand

$$D_k = S_k V_k$$

Utilisation law

$$U_k = X_k S_k = X D_k$$

Forced flow law

$$X_k = X V_k$$

Little's law

$$N = XR$$

Response time law

$$R = \frac{N}{X} - Z$$

Performance bounds

These are asymptotical bounds on throughput and response time that can be computed for different kinds of system.

Keep in mind that D is defined as:

$$D = \sum_{i=1}^{n} D_i$$

and $D_{\max}$ as:

$$D_{\max} = \max(D_i),\ \forall i$$

Throughput

System type	Bound
Terminal	$\frac{N}{ND+Z} \leq X(N) \leq \min\left(\frac{N}{D+Z}, \frac{1}{D_{\max}}\right)$
Batch	$\frac{1}{D} \leq X(N) \leq \min\left(\frac{N}{D}, \frac{1}{D_{\max}}\right)$
Transaction	$X(\lambda) \leq \frac{1}{D_{\max}}$

Response time

System type	Bound
Terminal	$\max(D, ND_{\max} - Z) \leq R(N) \leq ND$
Batch	$\max(D, ND_{\max}) \leq R(N) \leq ND$
Transaction	$D \leq R(\lambda)$

Turning point for system load We distinguish between low and high load in a system when its population is respectively lower or greater than:

$$N^* = \frac{D+Z}{D_{\max}}$$